

PREDICTING BIOLOGICAL NETWORKS AND THEIR EVOLUTIONARY DESIGN

Wiggins, Chris H.^{1,2}

¹Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY; ²Center for Computational Biology and Bioinformatics, New York, NY

Keywords: Machine Learning, Networks, Motifs, Transcriptional Regulation

Both the 'reverse engineering' of biological networks (for example, by integrating sequence data and expression data) and the analysis of their underlying design (by revealing the evolutionary mechanisms responsible for the resulting topologies) can be re-cast as problems in classification: predicting a categorical label in high-dimensional feature spaces. In the case of inferring biological networks, predicting up- or down- regulation of genes allows us to learn about the transcription factor binding sites (or 'motifs') and to generate a predictive model of transcriptional regulation. In the case of revealing evolutionary designs, quantitative, unambiguous model validation can be performed, clarifying which of several possible theoretical models of how biological networks evolve might best (or worst) describe real-world networks. In either case, by taking a machine learning approach, we statistically validate the models both on held-out data and via randomizations of the original dataset to assess statistical significance. By allowing the data to decide which features are the most important (based on predictive power rather than overabundance relative to an assumed null model) we learn models which are both statistically validated and biologically interpretable.

References:

- 1) Manuel Middendorf, Anshul Kundaje, Chris Wiggins, Yoav Freund, and Christina Leslie. Predicting genetic regulatory response using classification. ISMB 2004; q-bio/0411028
- 2) Manuel Middendorf, Anshul Kundaje, Mihir Shah, Yoav Freund, Chris H. Wiggins, and Christina Leslie. Motif discovery through predictive modeling of gene regulation. RECOMB 2005.
- 3) M. Middendorf, E. Ziv, and C. H. Wiggins. Inferring network mechanisms: the drosophila melanogaster protein interaction network. PNAS 2005; q-bio/0408010.
- 4) Manuel Middendorf, et al. Discriminative topological features reveal biological network mechanisms. BMC Bioinformatics 2004; q-bio/0402017.

These studies were supported in part by NSF ECS-0332479 (PI:CW), NSF ECS-0425850 (PI:CW), and 1U54CA121852-01A1 (National Center for the Multiscale Analysis of Genomic and Cellular Networks :MAGNet PI:Califano).